



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Authentication planning for XOR network coding

Katia Jaffrès-Runser — Cédric Lauradoux

N° 7562

March 2011



*Rapport
de recherche*



Authentication planning for XOR network coding

Katia Jaffrès-Runser ^{*}, Cédric Lauradoux ^{*}

Thème : Sécurité, Réseau, Codage en réseau
Équipes-Projets SWING

Rapport de recherche n° 7562 — March 2011 — 17 pages

Abstract: This paper formulates the authentication planning problem when network coding is implemented in a wireless sensor network. The planning problem aims at minimizing the energy consumed by the security application which is guaranteed using message authentication codes. This paper proposes a binary non-linear optimization formulation for this planning problem whose decision variables are the authentication decision of the nodes and the MAC modes of operation. It is illustrated for a butterfly topology. Results show that there is a real trade-off between energy efficiency and message throughput in this context.

Key-words: Wireless sensor networks, network coding, security, message authentication codes, planning, optimization

This paper has been submitted to IEEE NetCod 2011

^{*} Équipe SWING

Planification pour la sécurisation d'un codage en réseau

Résumé : Les réseaux sans-fil sont particulièrement vulnérables aux attaques par pollution dans lesquelles un attaquant externe est capable d'envoyer ces propres messages sur le réseau. Pour pouvoir détecter de telles attaques à la destination, un code d'authentification (MAC: Message Authentication Code en anglais) est rajouté à chaque paquet. Un noeud intermédiaire peut vérifier la validité d'un paquet de façon à limiter la portée de transmission d'un paquet pollué dans le réseau. Dans le cadre d'un réseau fortement contraint en énergie tel qu'un réseau de capteurs, le problème du déploiement d'une stratégie de sécurisation du réseau par MAC se pose. En effet, la consommation énergétique du réseau sera fortement influencée d'une part par le type de MAC utilisé dans le réseau et d'autre part par le choix des relais du réseau qui vérifieront le code des paquets avant de les retransmettre. Nous nous intéressons plus particulièrement au cas où le réseau de capteurs utilise une transmission par codage réseau de par sa plus grande vulnérabilité aux attaques par pollution. Ce type de réseau nécessite l'emploi de MAC dédiés (linéaires).

Dans ces travaux, nous proposons une formulation combinatoire du problème de planification de la sécurité. Dans cette formulation, nous minimisons l'énergie totale consommée par le réseau sécurisé pour la transmission d'un paquet par source dans le réseau. Les variables sont les décisions d'authentification binaires des noeuds. Nous illustrons ce modèle pour un réseau papillon pour lequel différentes distributions des probabilités d'attaque sur les liens sont considérées.

Mots-clés : Réseaux de capteurs, codage réseau, sécurité, message authentification codes, planification, optimisation

1 Introduction

Network coding [1, 2] are particularly vulnerable to *pollution attacks* [3] where an outsider adversary injects his malicious data. Indeed, network coding spreads the pollution by combining legitimate messages with polluted ones and therefore limiting the recovery probability of legitimate messages. Message authentication has to be ensured for end-to-end communications between any source and destination of the network. Pollution attacks can be defeated using message authentication codes (MACs). The primary goal of MAC is to prevent an adversary to tamper with the messages (substitution) and to forge its own messages (impersonation). A keyed cryptographic digest of the message, which can be ciphered or not, is appended to the message. The message is authenticated successfully if the destination is able to compute the same keyed signature than the one appended to the message, knowing the secret key shared with the emitter. A comprehensive survey of MAC can be found in [4]. In this paper, we make an extensive use of MAC based on universal hash functions (UHF-MAC) [5]. Such functions may exhibit linearity which is particularly suited for network coding and they have been used in past works [6–8] to thwart pollution attacks systematically by each node of the network.

In contrast to previous works [6–8], this paper addresses the problem of efficiently planning an authentication service for an energy constrained wireless network. A topical example is wireless sensor networking (WSN) whose security deployment has to guarantee low energy expenditure [9]. It is used as a case study herein. The security planning problem resumes to determining which nodes are going to authenticate the messages and which authentication strategies are the most energy efficient to deploy. We assume that the designer has some information on the distribution of the threat in the network. For instance, he may know that part of the network belongs to a trusted perimeter where security risks are low. Threat is modeled in this work with a probability of attack for each link of the network. A binary optimization formulation for the authentication planning problem is derived. Optimal solutions with respect to energy are provided and analyzed for a butterfly network topology with respect to various scenarios of attack.

The paper is structured as follows. Section 2 states the problem and Section 3 derives the according optimization model. Section 4 gives energy optimal roll out strategies for the butterfly network and Section 5 concludes the paper.

2 Authentication planning problem

In this paper, we only consider the case of XOR network coding [2] and not random linear network coding.

2.1 Attack topology

Pollution attacks are committed on the links of the network G . The number of links attacked and their location define an attack topology \mathcal{A} . We consider that the designer may not have a complete knowledge of the location of the attack at any point in time. Hence, he may have a confidence level in a link depending on its location in the network. For instance, links located inside of

Table 1: Energy performance of basic authentication strategies for a uniform attack topology.

| | Security maniac | Naïve strategy | Authentication planning |
|-------------------------|--------------------------------------|--|--|
| Authenticating Nodes | All relays + Destinations | Destinations only | Optimal selection w.r. threat + destinations |
| Objective | Detect threat asap | Limit unnecessary checks | Minimize energy |
| Low threat (small p) | Energy wasted for unnecessary checks | Quasi energy optimal | Energy optimal |
| High threat (high p) | Quasi energy optimal | Energy wasted for forwarding polluted messages | |

a trusted perimeter may have a higher confidence while the ones outside have a lower confidence. This feature is modeled using a probability of attack p_{ij} on a link (i, j) of the network which is defined as the probability of a message of being attacked on (i, j) .

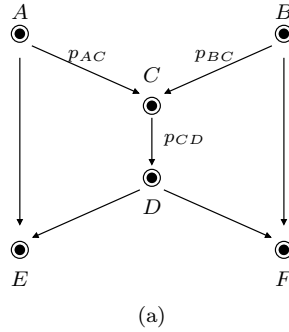


Figure 1: Butterfly network and attack topology.

An attack topology \mathcal{A} is defined by the distribution of the probabilities of attack for all the links of the network (see Fig. 1). For instance, this topology can be uniform and in this case, all links are attacked with the same probability p . Topology \mathcal{A} may as well model an attack localized on a single link (i, j) . In this case, \mathcal{A} is composed of a single non-null probability of attack p_{ij} .

2.2 Authentication strategies

The purpose of this paper is to derive a model that yields the *energy optimal authentication strategy* knowing an attack topology \mathcal{A} on a network G performing XOR network coding. An authentication strategy is defined by the subset of nodes that authenticate the messages and the modes of UHF-MAC used (presented later in this section).

First of all, it is important to note that the sources generate the messages with their corresponding digests and that the destinations always verify what they have received. Consequently, the destinations are always able to disregard

polluted messages. These checks are mandatory and their cost in energy is incompressible.

For all relaying nodes other than the destinations, we have a degree of freedom: they may or not authenticate the messages. This may incur a certain authentication cost in energy at the relays. As shown in [8] performing verification is as energy expensive as sending plus receiving a message. However, in some cases it may be beneficial to the overall network energy performance since polluted messages are not uselessly forwarded towards the destination. In the case of network coding, it will also prevent the creation of polluted combination and preserve the throughput.

In terms of security planning, two extreme strategies are often considered. On the one side, the *security maniac strategy* emphasizes on detecting an attack as soon as possible, limiting the pollution in the network. All the relaying nodes authenticate the messages as considered in [3, 6, 7]. Unnecessary verifications can lead to a waste of energy. On the other side, the *security naïve strategy* considers that an end-to-end authentication is sufficient and that there is no need to empower relays with authentication capabilities. Forwarding polluted messages incurs both a same energy waste and a throughput reduction.

Intermediate strategies are possible to improve the energy and throughput performance. Table 4 resumes all strategies. Finding optimal strategies is the aim of this work which can be achieved by solving an optimization problem. More specifically, we define the *authentication planning problem* that minimizes the overall energy consumption of a WSN knowing its topology, the network coding rules, the authentication MAC modes and the attack topology existing in the network.

2.3 MAC schemes

It has been established by Apavatjirut et al. in [8] that MACs based on the classical primitives that are block ciphers or hash functions imply an energetic cost too important for the relaying nodes of a WSN. The same observation holds for the underlying primitives (exponentiation) used in [7]. On the opposite, MACs based on UHFs [6, 8] offer more flexibility for the authentication if we use an ϵ -almost XOR universal hash (ϵ -AXU) function h is (see [5] for more details). The most interesting property for our problem is the linearity of these functions: $h(m_1) \oplus h(m_2) = h(m_1 \oplus m_2)$ with m_1 and m_2 two n -bit messages. We voluntarily skip the details related to this function as they are not essential to understand the core of this paper (see [6, 8] for further details).

Exploiting the linearity is particularly interesting for authentication in the context of network coding. Let us consider a node in the network who has to combine (XOR) ℓ messages and their corresponding authentication codes. A MAC based on ϵ -AXU function offers three possibilities for authentication: (i) the node authenticates each message individually, combines (XOR) the valid ones and computes the authentication code of this sum. Then, the message is forwarded. We refer to this mode of operation by AXF throughout the paper. The AXF mode requires ℓ verifications, *i.e.* ℓ computations of the MAC. (ii) The node checks that the sum modulo two of the authentication codes is equal to the authentication codes of the sum modulo two of the messages. By doing so, it exploits the linearity of the MAC to reduce the authentication to a single computation of the MAC. We refer to this mode of operation as XAF. The

drawback of the XAF is that the node forwards a message if and only if the ℓ incoming messages are not polluted. (iii) The node can also simply forward the sum modulo two of messages alongside the sum modulo two of the authentication codes. Any verification is delegated to other nodes. We refer to this mode of operation as XF.

3 Optimization model

The authentication planning problem is formulated in the following using a binary integer program.

3.1 Network model

We assume that the network topology is known. The network is modeled using a directed acyclic graph $G(\mathcal{V}, \mathcal{E})$ having vertex set \mathcal{V} and edge set $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$. Without loss of generality, $\mathcal{V} = \{1, 2, \dots, |\mathcal{V}|\}$. For each node $i \in \mathcal{V}$, $\vec{\mathcal{N}}_i$ and $\overleftarrow{\mathcal{N}}_i$ are the sets of edges leaving from and the set of edges going into i , respectively. Formally $\vec{\mathcal{N}}_i = \{(i, j) | (i, j) \in \mathcal{E}\}$ and $\overleftarrow{\mathcal{N}}_i = \{(j', i) | (j', i) \in \mathcal{E}\}$.

A set of sources \mathcal{O} and destinations \mathcal{D} is defined. Since we address the WSN case, \mathcal{O} is the set of sensors having data to report in multicast to the $|\mathcal{D}|$ sink nodes and we consider $|\mathcal{D}| \ll |\mathcal{O}|$. Source and destination nodes do not relay the information. As a consequence, the network we are modeling is composed of a set of relay nodes $\mathcal{R} = \mathcal{V} \setminus (\mathcal{O} \cup \mathcal{D})$. In the following, we consider that the number of relays in the network is $N = |\mathcal{R}|$. So far, we do not consider any propagation losses and assume a perfect channel transmission.

The vertices \mathcal{V} are partitioned into two groups of nodes: a subset $\mathcal{R}_c \in \mathcal{R}$ of relays performing XOR network coding and $\mathcal{R}_f = \mathcal{R} \setminus \mathcal{R}_c$ which are simply forwarding messages. Knowing $G(\mathcal{V}, \mathcal{E})$, simple rules are set to define \mathcal{R}_c and \mathcal{R}_f : the coding relays have more than one edge coming into them (i.e. $|\overleftarrow{\mathcal{N}}_i| > 1$) while forwarding-only relays \mathcal{R}_f are characterized by a single incoming edge (i.e. $|\overleftarrow{\mathcal{N}}_i| = 1$).

The binary quantity $c_i \in \{0, 1\}$ is fixed to differentiate nodes of \mathcal{R}_c performing network coding from nodes of \mathcal{R}_f that are simply forwarding messages. Hence, for any node $i \in \mathcal{R}_c$, $c_i = 1$ and for $i \in \mathcal{R}_f$, $c_i = 0$.

Attack topology Security threats are modeled in G by a valuation p_{ij} on edge $(i, j) \in \mathcal{E}$. The set of all valuations $\mathcal{A} = \{p_{ij} | (i, j) \in \mathcal{E}\}$ defines an attack topology. We assume that the attacks are independent on the edges of the graph. The attack topology is considered as being known by the network designer. In this setup, energy optimal security strategies can be derived as shown in the following.

The following optimization model captures the impact of the previously described MAC strategies on the overall energy consumption of the network.

3.2 Optimization variables

Authenticate variable Any node in \mathcal{R} may or may not authenticate messages, whether this node is a coding or a forwarding-only node. Let $x_i \in \{0, 1\}$

be the first main binary variable of this model. If $x_i = 1$, node i authenticates each incoming message while if $x_i = 0$, it never authenticates.

If a node authenticates a message (i.e. $x_i = 1$), it has two opposite effects on the overall energy consumption:

- more energy is spent by the authentication process,
- but energy for forwarding polluted messages is saved.

MAC mode of operation As shown earlier, XAF and AXF modes do not yield the same energy consumption. Let $m_i \in \{0, 1\}$ be a binary variable that gives the mode of authentication used by a coding node $i \in \mathcal{R}_c$. If $m_i = 1$, we have XAF and if $m_i = 0$ we have AXF. This variable can be interpreted as whether the XOR operation is done before authentication ($m_i = 1$; XAF mode) or after authentication ($m_i = 0$; AXF mode).

Node and network authentication strategy The *node authentication strategy* S_i is defined for any coding node $i \in \mathcal{R}_c$ by the tuple $S_i = (x_i, m_i, c_i)$ and for any forwarding-only node $i \in \mathcal{R}_f$ by the pair $S_i = (x_i, c_i)$. Table 2 gives the correspondence between the possible tuples and the MAC modes of operation for any coding and forwarding-only node. For forwarding-only nodes, the value of m_i is undetermined since no XOR step is performed.

The *network authentication strategy* ξ is defined by the set of node authentication strategies for all nodes of the network $\xi = \{S_i, \forall i \in \mathcal{R}\}$.

Table 2: MAC modes and corresponding security strategies

| Node security strategies | | |
|--------------------------|-----|-------------------------------|
| Coding node | AXF | $(x_i = 1, m_i = 0, c_i = 1)$ |
| | XAF | $(x_i = 1, m_i = 1, c_i = 1)$ |
| | XF | $(x_i = 0, m_i = 1, c_i = 1)$ |
| Relay node | AF | $(x_i = 1, c_i = 0)$ |
| | F | $(x_i = 0, c_i = 0)$ |

3.3 Forwarding decisions

We define the *forwarding decision of a node i* as the probability that this node decides to transmit a received message. Let $f_i \in [0, 1]$ be the forwarding probability of node i . If authentication is performed by node i , this decision is positive if no polluted message is detected. This decision is a direct consequence of the node authentication strategy $S_i(x_i, m_i)$ and the probability of a polluted message to arrive from a direct neighbor node k to node i .

Let define \mathbf{P}_{ki} as the *probability of a polluted message to arrive at node i coming from node k* . This probability is a function of node authentication strategy S_k , the forwarding decisions and the attack probabilities related to all the paths between the sources $s \in \mathcal{O}$ and i going through k . Its derivation is given after the forwarding decision description.

The forwarding decision f_i of node i depends on its node type c_i . The global formulation of f_i is:

$$f_i = (1 - c_i) \cdot f_i^R + c_i \cdot f_i^C \quad (1)$$

where f_i^R and f_i^C are the forwarding probabilities for the case node i is a forwarding or a coding node, respectively.

Forwarding node For $i \in \mathcal{R}_f$ ($c_i = 0$), the forwarding decision is function of x_i and \mathbf{P}_{ki} . If $a_i = 0$, the node simply forwards every received message and hence $f_i = 1$. Else ($x_i = 1$) the node forwards with probability $f_i = (1 - \mathbf{P}_{ki})$ which represents the probability of an unpolluted message to arrive in i . The forwarding probability for the case $c_i = 0$ is derived as:

$$f_i^R = (1 - x_i) + x_i(1 - \mathbf{P}_{ki}) \quad (2)$$

Coding node For $i \in \mathcal{R}_c$ ($c_i = 1$), the forwarding decision depends on the MAC strategy. For the case of AXF, a node forwards a message if at least one of its incoming messages is non-polluted which happens with probability $1 - \prod_{k \in \mathcal{N}_i} \mathbf{P}_{ki}$. For the case of XAF, a message is forwarded if all messages XOR-ed together are non-polluted which happens with probability $\prod_{k \in \mathcal{N}_i} (1 - \mathbf{P}_{ki})$.

A closed-form derivation of the forwarding decision for any type of node of the network is given by:

$$f_i^C = (1 - x_i) + x_i \left[m_i \cdot \prod_{k \in \mathcal{N}_i} (1 - \mathbf{P}_{ki}) + (1 - m_i) \left(1 - \prod_{k \in \mathcal{N}_i} \mathbf{P}_{ki} \right) \right] \quad (3)$$

The pollution probability \mathbf{P}_{ki} A message coming into node i from a neighbor node can be polluted for two reasons: *i*) node k sends a message that is not polluted and the message gets polluted on the link between k and i following the local probability of attack p_{ki} on (k, i) ; *ii*) node k forwards a message that is polluted (this is only the case if node k does not implement an authentication function, i.e. $x_k = 0$).

For the case $x_k = 1$, node k authenticates and the pollution probability is equal to $\mathbf{P}_{ki} = f_k \cdot p_{ki}$, which is the probability that node k forwards an unpolluted message and that it can only be polluted by an attack on link (k, i) .

For the case $x_k = 0$, node k cannot detect if it forwards a polluted message or not. Hence, the probability for the message sent by k to be polluted depends on the previous history of the message in the network. Hence, it is derived recursively knowing the values of the forwarding and attack probabilities on all paths coming into node k . In this case, $\mathbf{P}_{ki} = 1 - (1 - p_{ki}) \cdot \prod_{l \in \mathcal{N}_k} (1 - \mathbf{P}_{lk})$, where $\prod_{l \in \mathcal{N}_k} (1 - \mathbf{P}_{lk})$ is the probability for a message to arrive in k without being polluted on the links coming into k .

A global formulation of \mathbf{P}_{ki} with respect to x_k is given by:

$$\mathbf{P}_{ki} = f_k [x_k \cdot p_{ki} + (1 - x_k) \cdot \left(1 - (1 - p_{ki}) \cdot \prod_{l \in \mathcal{N}_k} (1 - \mathbf{P}_{lk}) \right)] \quad (4)$$

For the case $k = s \in \mathcal{O}$, $\mathbf{P}_{si} = p_{si}$ since $f_s = 1$.

Pollution and forwarding probability Since the network is a directed acyclic graph, there are no loops in the network and the values of the pollution and forwarding probabilities exist and can be derived for any node of the network. The causal dependency between the definitions of these probabilities is rooted in the network coding of the messages originating from different paths. In order to compute f_i for node i , the pollution probabilities on all its incoming links $\{\mathbf{P}_{ki}, \forall k \in \overleftarrow{\mathcal{N}}_i\}$ are needed. These values depend on the forwarding probabilities of the intermediary nodes that belong to the existing paths joining the source nodes to i .

We consider first the case of a layered network where the network is divided into layers of nodes. The sources are connected to nodes of layer one but not to nodes of layer 2, nodes of layer 1 are connected to nodes of layer 2 but not to nodes of layer 3, etc. In this case, pollution and forwarding probabilities can be computed layer by layer. For layer 1 nodes, $\mathbf{P}_{si} = p_{si}$ and f_i is deduced using (1). Then, \mathbf{P}_{ij} is computed according to (4) for layer 2 nodes and f_j is derived according to (1) for layer 2 nodes as well. The process is repeated until the destination layer is reached. This iterative algorithm can be extended to support the case of a more general DAG, but for conciseness purposes, it is not presented herein.

3.4 Energy cost function

The energy cost function $\mathcal{F}_E(\xi)$ counts the energy spent for the end-to-end transmission of one message sent by the sources $s \in \mathcal{O}$ to their destinations \mathcal{D}_s for a specific network authentication strategy (or solution) ξ :

$$\mathcal{F}_E(\xi) = \mathcal{F}_{\mathcal{O}}(\xi) + \mathcal{F}_{\mathcal{R}}(\xi) + \mathcal{F}_{\mathcal{D}}(\xi) \quad (5)$$

where $\mathcal{F}_{\mathcal{O}}(\xi)$, $\mathcal{F}_{\mathcal{R}}(\xi)$ and $\mathcal{F}_{\mathcal{D}}(\xi)$ are the costs in energy relative to the energy expenditure of source nodes, relay nodes and destination nodes, respectively.

Table 3: Energy costs ($\times 10^{-4} J$) for the atomic actions. Values are given for the transmission of one message.

| | | |
|--------------------------|-----------|------------|
| Emission | Q_T | 0.556851 |
| Reception | Q_R | 0.7995405 |
| Authentication (UHF-MAC) | Q_A | 1.686154 |
| XOR of 2 messages | Q_{XOR} | 0.00003135 |

The costs in energy for the atomic actions are listed in Table 3. They have been collected in [8] using the WSim/eSimu energy estimation tool [10] for a TI MSP430 based platform and a Ti CC2420, 802.15.4 compliant, radio device similar to TelosB nodes. It is worth mentioning that one authentication is as expensive as a combined message emission and reception.

The cost related to the transmission of a message by the source nodes is directly proportional to the number of sources $\mathcal{F}_{\mathcal{O}} = |\mathcal{O}| \cdot (Q_T + Q_A)$. The cost related to the reception of a message by the destination nodes depends

on the number of messages N_d destination d will receive from its previous hop neighbors:

$$\mathcal{F}_{\mathcal{D}} = \sum_{d \in \mathcal{D}} N_d \cdot (Q_R + Q_A)$$

where $N_i = \sum_{k \in \mathcal{N}_i} f_k$.

The derivation of the energy consumption for all relays in the network is given by:

$$\mathcal{F}_{\mathcal{R}} = \sum_{i \in \mathcal{R}} [N_i \cdot Q_R + x_i \cdot \mathcal{F}_A(c_i, m_i) + f_i \cdot Q_T]$$

The function $\mathcal{F}_A(c_i, m_i)$ gives the energy consumption of authentication with respect to the type of node (coding or forwarding) and the MAC mode considered. It is defined by:

$$\begin{aligned} \mathcal{F}_A(c_i, m_i) = & c_i \cdot [Q_{XOR} \cdot (N_i - 1) + \\ & m_i \cdot Q_{XAF}(i) + (1 - m_i) \cdot Q_{AXF}(i)] + (1 - c_i)Q_A \cdot \mathbf{P}_i^{Rec} \end{aligned}$$

where $Q_{AXF}(i)$ and $Q_{XAF}(i)$ are the costs for authenticating a message using AXF and XAF, respectively. $\mathbf{P}_i^{Rec} = 1 - \prod_{k \in \mathcal{N}_i} (1 - f_k)$ is the probability of node i to receive at least one message from its one hop neighbors.

XAF has the cost of authenticating a single message ($Q_{XAF}(i) = Q_A \cdot \mathbf{P}_i^{Rec}$) since it is performed on the XOR-ed version of the incoming messages. $Q_{AXF}(i)$ is a function of the number of messages received at node i since each incoming message is authenticated individually. It is derived as $Q_{AXF}(i) = Q_A \cdot N_i$

3.5 Optimization problem definition

We recall that \mathcal{R} is a set of N relays of $c = |\mathcal{R}_c|$ coding and $N - c$ forwarding relays. The *security planning problem* can be formulated by the binary integer program as follows.

$$\begin{aligned}
& \min && \mathcal{F}_E(\xi) \\
s.t. & f_i = && (1 - c_i) \cdot f_i^R + c_i \cdot f_i^C, && \forall i \in \mathcal{R} \\
& f_i^R = && (1 - x_i) + x_i(1 - \mathbf{P}_{ki}), && \forall i \in \mathcal{R}_f \\
& f_i^C = && (1 - x_i) + x_i \left[m_i \cdot \prod_{k \in \bar{\mathcal{N}}_i} (1 - \mathbf{P}_{ki}) + \right. \\
& && \left. (1 - m_i)(1 - \prod_{k \in \bar{\mathcal{N}}_i} \mathbf{P}_{ki}) \right], && \forall i \in \mathcal{R}_c \\
& \mathbf{P}_{ki} = && f_k [x_k \cdot p_{ki} + \\
& && (1 - x_k) \cdot \left(1 - (1 - p_{ki}) \cdot \prod_{l \in \bar{\mathcal{N}}_k} (1 - \mathbf{P}_{lk}) \right)] \\
& \xi = && \{S_i, \forall i \in \mathcal{R}\} \\
& S_i = && \begin{cases} (x_i, m_i, c_i) & \forall i \in \mathcal{R}_c \\ (x_i, c_i) & \forall i \in \mathcal{R}_f \end{cases} \\
& (c_i, x_i) \in && \{0, 1\} \times \{0, 1\} && \forall i \in \mathcal{R} \\
& m_i \in && \{0, 1\} && \forall i \in \mathcal{R}_c \\
& p_{ki} \in && [0, 1] && \forall (i, j) \in \mathcal{E}
\end{aligned}$$

The solution set has a cardinality of $3^c \cdot 2^{N-c}$. The energy cost is not linear and hence, the problem is not an binary linear program.

Table 4: Description of the six solutions for the butterfly network

| Strategy | (x_C, m_C) | x_D | Description |
|------------|--------------|-------|--------------------------------|
| (XF ; F) | $(0, -)$ | 0 | C, D forward only |
| (XF ; AF) | $(0, -)$ | 1 | Authentication on D only |
| (AXF ; F) | $(1, 0)$ | 0 | AXF on C only |
| (XAF ; F) | $(1, 1)$ | 0 | XAF on C only |
| (AXF ; AF) | $(1, 0)$ | 1 | AXF on C , Auth. on D only |
| (XAF ; F) | $(1, 1)$ | 1 | XAF on C , Auth. on D |

4 Results

The security planning problem is illustrated for the butterfly network (cf. Fig. 1). C is a coding node and D a forwarding relay. The six network authentication strategies are given in Table 4. For instance, the first strategy (XF ; F) is the security naïve strategy where neither C nor D are authenticating. On the opposite, the two last entries are security maniac strategies where both nodes are authenticating.

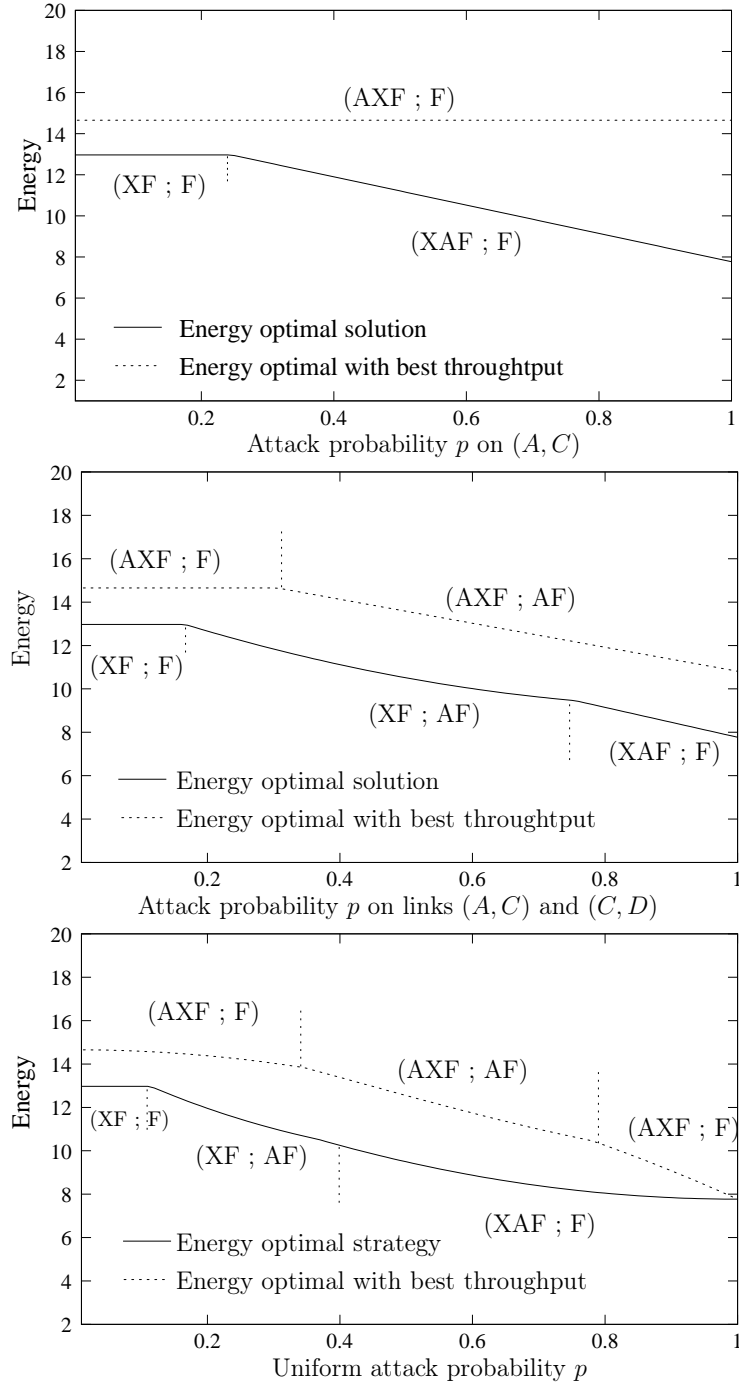


Figure 2: Energy optimal strategy with and without best throughput constraint with respect to attack probability p when 1, 2 and 3 links are attacked.

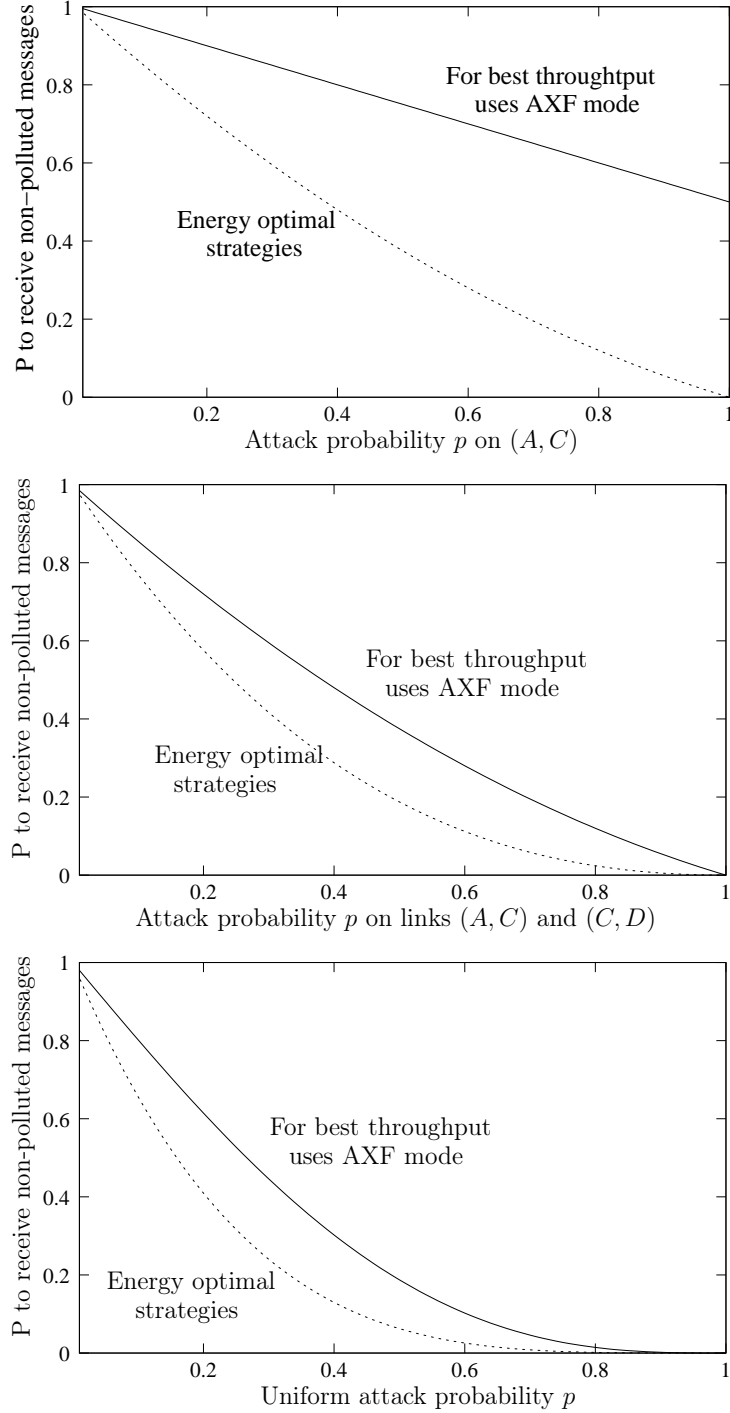


Figure 3: Average probability to receive non-polluted messages for energy optimal strategies with and without best throughput constraint with respect to attack probability p when 1, 2 and 3 links are attacked.

In this analysis, we only consider attack topologies that involve the links (A, C) , (B, C) and (C, D) since attacks on other links can only be detected by destinations E and F . We describe three scenarios of attack:

- The attack targets a *single link*. Since the network is symmetrical, only the cases of an attack on links (A, C) and (C, D) are relevant. We chose to show the results for an attack on link (A, C) .
- The attack targets *two links*. Again, only the cases where attacks are on the pair of links $(A, C)/(B, C)$ and $(A, C)/(C, D)$ can be considered for symmetry purposes. The results related to the $(A, C)/(C, D)$ pairs are presented here. We assume that the attacks on both links arise with the same probability (i.e. $p_{AC} = p_{CD} = p$).
- The attack targets the *incoming links of the relaying nodes*. In this case, all three links are attacked with the same uniform probability (i.e. $p_{AC} = p_{BC} = p_{CD} = p$).

The results related to the energy optimal strategies are given on Fig. 2 and Fig. 3. Fig. 2 presents the total energy spent $\mathcal{F}_E(\xi^*)$ by the optimal strategy ξ^* with respect to the attack probability p on the link(s) targeted by the attacker. Two cases are considered. In the first one, we look for the *energy optimal strategy* which minimizes total energy following the problem defined in Section 3. In the second case, we show the performance of the *energy optimal with best throughput strategy* which looks for the strategy that maximizes the throughput at minimal energy.

Throughput is measured in our case by the average probability P_{th} for both destinations E and F to decode the messages of A and B . Destinations can decode messages from A and B if they are non-polluted. If authentication is performed at coding node C , this probability depends on the MAC mode. For AXF, this probability is higher than for XAF because messages that are not polluted are always forwarded. A general expression for the butterfly network is $P_{th} = 0.5 \cdot f_C \cdot (1 - p_{CD}) [2 - p_{AC} - p_{BC}]$, where $f_C = 1 - p_{AC} \cdot p_{BC}$ for AXF and $f_C = (1 - p_{AC})(1 - p_{BC})$. Fig. 3 gives the values of P_{th} with respect to the attack probability p for the *energy optimal* and *energy optimal with best throughput* strategies.

Results show that strategies that minimize the number of forwarded messages are the most energy efficient ones. AXF never belongs to an *energy optimal strategy* because the energy cost of verification is high but also because it transmits more messages. However, this is the only MAC mode that mitigates the spread of pollution induced by network coding. As a first conclusion, if throughput guarantee is the main concern, energy has to be spent for authentication by using AXF.

When the probability of attack is low, the strategy where C and D simply forward messages is the most energy efficient since there are no security checks. However, when p increases, more energy can be saved by authenticating messages at the relays. For instance, for a single attack on (A, C) and $p > 0.24$, strategy (XAF;F) saves more energy because D does not relay combined messages that are polluted. For high p , XAF mode on C provides the minimum energy but drastically reduces throughput. Consequently, for small p , the *energy optimal strategy* is to be favored because the loss of throughput is less important

while for higher p , more energy-consuming modes of MAC have to be considered to favor throughput.

5 Conclusions and Perspectives

This paper formulates the energy efficient authentication planning problem for XOR network coding. It formulates the problem as a binary non-linear optimization problem that minimizes the overall energy consumption of the secured network. Optimal roll-out of security-enabled nodes can be deduced together with their appropriate MAC mode. Results for the butterfly topology exhibit the trade-off between energy efficiency and throughput of non-polluted messages as a function of the MAC mode considered.

Determining optimal security roll-outs should now be done for larger networks. In this context, exhaustive search is not scalable and proper optimization tools need to be derived in future works. A relaxed version of the problem can be formulated where the binary authentication variable becomes a message authentication probability. In this case, we move from a hard decision to a soft decision model, which could be easier to solve. A multiobjective optimization approach could be considered as well in order to find the Pareto bound maximizing network throughput and minimizing energy expenditure. This work opens several perspectives as well. Extending this study to the case of random linear network coding is important for future work. Similarly, the definition of distributed algorithms that converge to energy efficient authentication strategies is a practical result of interest.

References

- [1] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, 2000.
- [2] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Médard, and J. Crowcroft, "XORs in the air: practical wireless network coding," *IEEE/ACM Transactions on Networking*, vol. 16, no. 3, pp. 497–510, 2008.
- [3] J. Dong, R. Curtmola, and C. Nita-Rotaru, "Secure network coding for wireless mesh networks: Threats, challenges, and directions," *Computer Communication*, vol. 32, no. 17, pp. 1790–1801, 2009.
- [4] A. J. Menezes, S. A. Vanstone, and P. C. Van Oorschot, *Handbook of Applied Cryptography*. CRC Press, Inc., 1996.
- [5] H. Krawczyk, "LFSR-based Hashing and Authentication," in *Advances in Cryptology - CRYPTO '94*, ser. Lecture Notes in Computer Science 839. Springer-Verlag, 1994, pp. 129–139.
- [6] S. Agrawal and D. Boneh, "Homomorphic MACs: MAC-Based Integrity for Network Coding," in *Applied Cryptography and Network Security - ACNS 2009*, ser. LNCS 5536, 2009, pp. 292–305.

- [7] D. Boneh, D. Freeman, J. Katz, and B. Waters, “Signing a linear subspace: Signature schemes for network coding,” in *Public Key Cryptography - PKC 2009*, ser. Lecture Notes in Computer Science 5443. Springer Verlag, 2009, pp. 68–87.
- [8] A. Apavatjrut, W. Znaidi, A. Fraboulet, C. Goursaud, C. Lauradoux, and M. Minier, “Energy friendly integrity for network coding in wireless sensor networks,” in *International Conference on Network and System Security - NSS 2010*, September 2010, pp. 1–8.
- [9] A. Perrig, J. A. Stankovic, and D. Wagner, “Security in wireless sensor networks,” *Communication of the ACM*, vol. 47, no. 6, pp. 53–57, 2004.
- [10] N. Fournel, A. Fraboulet, and P. Feautrier, “Embedded software energy characterization: Using non-intrusive measures for application source code annotation,” *Journal of Embedded Computing*, vol. 3, no. 3, 2009.

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 3 |
| 2 | Authentication planning problem | 3 |
| 2.1 | Attack topology | 3 |
| 2.2 | Authentication strategies | 4 |
| 2.3 | MAC schemes | 5 |
| 3 | Optimization model | 6 |
| 3.1 | Network model | 6 |
| 3.2 | Optimization variables | 6 |
| 3.3 | Forwarding decisions | 7 |
| 3.4 | Energy cost function | 9 |
| 3.5 | Optimization problem definition | 10 |
| 4 | Results | 11 |
| 5 | Conclusions and Perspectives | 15 |



Centre de recherche INRIA Grenoble – Rhône-Alpes
655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
<http://www.inria.fr>
ISSN 0249-6399